様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２７ 年　 ６ 月　 １ 日現在

機関番号：３２６６０
研究種目：研究活動スタート支援
研究期間：2013～2014
課題番号：２５８８００１７

研究課題名（和文）高次元データに対する正規性検定および多変量モデルの選択

研究課題名（英文）Normality test and celection of multivariate model for high-dimension data

研究代表者
榎本　理恵（Enomoto, Rie）

東京理科大学・理学部・助教

研究者番号：３０７１１７６７

交付決定額（研究期間全体）：（直接経費）　　1,900,000円

研究成果の概要（和文）：本研究課題では、正規性検定統計量と成長曲線モデルに対する規準量について議論した。成長曲線モデルに関しては高次元データのもと、標本数とグループ数がともに大きいもとでの規準量を新たに提案した。それぞれの仮定のもとでの漸近分布と各準量について一致性を議論した。本研究結果は数値計算を行って有用性を確認した。

研究成果の概要（英文）：Our studies are normality test statistics and information criteria for growth curve model. For growth curve model, we proposed the information criteria under high-dimension or under both group and sample size is large. We discussed consistency properties many information criteria, respectively. Our results are checked numerically by conducting a Monte Carlo simulation.

研究分野：多変量解析

キーワード：多変量解析　高次元データ

１．研究開始当初の背景

（１）さまざまな統計モデルや分析手法は“正規分布”を仮定していることが多い．それは正規分布が取り扱いし易く，標本数が十分に確保できるとき，データは漸近的に正規分布に従うと仮定できるという理由からである．しかし，実際に分析するデータが正規分布であるかどうかを吟味する分析者は多くはない．多変量の場合，一変量と異なりQ-Qプロットを描き視覚的な判断や，特性関数に基づく統計量が使い手にとって取扱いし難いといった背景がある．そこで使い手が容易に検定できるよう積率に基づく検定方法が従来から提案されている．大標本理論に対して，Mardia (1970)が定義した歪度と尖度がよく知られており，これは三次と四次積率に基づいて定義されている．Mardia (1970)による定義は，主成分スコアのみを用いたSrivastava (1984)の定義と異なり，多くのデータを用いて総合的に判断ができる点，データが正規分布でないときに正しく判断する検出力が高いことが利点とされている．

（２）成長曲線モデルというのは，例えば，女子と男子の成長（身長）が１次直線，2次直線と予測できるようなモデル化のことである．よって，各個体の時間変化に対して適当な曲線を当てはめ，更に各個体の繰り返し測定に未知の分散共分散行列をもつ多変量正規性などを想定するモデルである．AIC規準は真（正解）のモデルを含む多項式モデルを正しく選択する規準として古くから使われている．Satoh, Kobayashi and Fujikoshi (1997)は成長曲線モデルに対する大標本理論の下での規準量を提案しているが，高次元データに適用した場合，誤ったモデルを選択することを確認している．

２．研究の目的

本研究課題は，高次元データに対する正規性検定と多項式モデルにおける変数選択規準の提案を主とする．申請者はこれまでに，標本数が次元数（説明変数）よりも大きい大標本理論の下で，データが正規分布に従うかどうかを調べる正規性検定について積率に基づいた統計量の提案と実用性を確認している．本研究課題では，医学や薬学で現れるような標本数よりも次元数が大きい高次元データを想定している．また，各個体について経時的に繰り返し測定されるデータのモデル化の一つである成長曲線モデルに対して，データから適切なモデルを選ぶための変数選択規準の構成を行う．基礎研究を基に関連研究の体系的な整理・理解から，既存の手法が破綻する状況の確認と手法や統計量の改良を目指す．

３．研究の方法

（１）高次元データに対する正規性検定統計量を提案することは容易ではない．そこでまずは大標本理論の下で統計量の性質を吟味することから始める．正規性検定に対する統計量はさまざま提案されているが，その中で簡便な方法として考えられているのが積率に基づく統計量である．本研究課題では簡便さに注目し，積率に基づく統計量とその他の統計量との比較を行う．主に正規分布の下での近似精度と検出力の観点から比較をする．検出力とは得られたデータが正規分布に従っていないときに正しく判断を行うことであり，検出力が高いほどより精度の良い統計量と言える．想定する分布は多種に渡り，また，多くのパラメータに対して実験を行うことになる．

（２）成長曲線モデルに対する変数選択規準量と分布について議論する．想定する漸近枠組みとしては「標本数と次元数がともに大きい」及び「標本数とグループ数がともに大きい」である．候補のモデルの中に真のモデルが含まれる場合でのAIC規準の構成，モデルの選択確率を理論的に導出

し，さらに一致性についても議論を行う．「一致性がある」とは例えば大標本理論の下で，標本数を増加させた場合に真のモデルを選ぶ確率が１になることである．尤度に基づいた AIC 規準量は大標本理論の下で一致性がないことが良く知られている．提案されているさまざまな規準量については，それぞれ性能の比較をすることが難しいことが問題点とされている．そこで規準量の一致性（特性）を議論することは重要である．本研究の性質上，最終的に数値実験により導出した統計量および AIC 規準の正当性・妥当性を確認する．

４．研究成果

（１）ルブリン生命科学大学の Hanusz 教授との共同研究については，尖度と歪度を用いた正規性検定統計量を中心に主要な統計量についてさまざまな分布に関して特性を吟味した学術論文を投稿中である．検出力については，よく知られている分布として多変量 t 分布や多変量一様分布などさまざまな分布の仮定を置いている．正規性検定統計量についての検出力比較はそれほど研究されていない．また，積率に基づいた統計量についても，近似精度が改良された統計量なども比較対象に含めている．

（２）成長曲線モデルに関する研究結果は以下の通りである．

①高次元枠組みの下で規準量の提案

②別の漸近枠組みの下での規準量の提案

③②の下での各規準量の推測

(②，③では標本数とグループ数が両方大きいという仮定を置いている．)

以上３つの内容については，それぞれ一致性を議論している．①では大標本理論の下で提案されている規準量は漸近枠組みが崩れた場合には有用でないこと，高次元枠組みの下で提案した規準量は大標本理論及び高次元枠組みの下，両方で有用であることを確認している．

②では，①において提案した規準量は高次元枠組みの下で一致性をもたないことを受け，異なる漸近枠組みの下で一致性をもつことを示した．また，仮定した漸近枠組みに対する規準量の分布についても議論している．

③で仮定した漸近枠組みの下では大標本枠組みの下では一致性をもつ規準量について，一致性がないことなどを確認している．また，尤度に基づいた規準量および Cp 型や修正した規準量についての漸近分布，さらに一致性についても議論している．

全ての結果においては数値実験で精度を確認しており，それぞれ学術論文に掲載が決定している．

＜引用文献＞

① Satoh, K. , Kobayashi, M. and Fujikoshi, Y. (1997). “ Variable selection for the growth curve model”. Journal of Multivariate Analysis, 60, 277－292.

② Mardia, K. V. (1970). “Measures of multivariate skewness and kurtosis with applications”, Biometrika, 57, 519－530.

③ Srivastava, M. S. (1984). “A measure of skewness and kurtosis and a graphical method for assessing multivariate normality”, Statistics & Probability Letters, 2, 263－267.

５．主な発表論文等

（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計　３　件）

① Enomoto, R., Sakurai, T. and Fujikoshi, Y. (2014). “Consistency properties of AIC, BIC, Cp and their modifications in the growth curve model under a

large-(q; n) framework”, SUT Journal of Mathematics, submmitted.（査読有）
http://www3.ma.kagu.tus.ac.jp/sutjmath/

② Enomoto, R., Sakurai, T. and Fujikoshi, Y. (2013). “Consistency of AIC and its modification in the growth curve model under a large-(q, n) framework”, SUT Journal of Mathematics, 49, 93－107.（査読有）
http://www3.ma.kagu.tus.ac.jp/sutjmath/

③ Fujikoshi, Y., Enomoto, R. and Sakurai, T. (2013). “High-dimensional AIC in the growth curve model”, Journal of Multivariate Analysis, 122, 239－250.（査読有）
DOI：10.1016/j.jmva.2013.07.006

〔学会発表〕（計　3　件）

① 榎本 理恵, 櫻井 哲朗, 藤越 康祝, 成長曲線モデルにおける各規準量の高次元一致性, 統計関連学会連合大会 2014, 2014 年 9 月 14 日, 東京大学　本郷キャンパス（東京都・文京区）

② 榎本 理恵, 櫻井 哲朗, 藤越 康祝, 成長曲線モデルにおける各規準量の高次元漸近分布, 統計関連学会連合大会 2013, 2013 年 9 月 9 日, 大阪大学　豊中キャンパス（大阪府・豊中市）

③ 榎本 理恵, 岡本 直也, The omnibus test statistics for normality based on sample moments, 統計関連学会連合大会 2013, 2013 年 9 月 9 日, 大阪大学　豊中キャンパス（大阪府・豊中市）

６．研究組織

（１）研究代表者

榎本　理恵（ENOMOTO, Rie）
東京理科大学・理学部・数理情報科学科・助教
研究者番号：３０７１１７６７